Optimal Electric Vehicle Charging Strategy with Markov Decision Process and Reinforcement Learning Technique

Ding, Tao; Zeng, Ziyu; Bai, Jiawen; Qin, Boyu; Yang, Yongheng; Shahidehpour, Mohammad

Published in:
I E E E Transactions on Industry Applications

DOI (link to publication from Publisher):
10.1109/TIA.2020.2990096

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

Optimal Electric Vehicle Charging Strategy with Markov Decision Process and Reinforcement Learning Technique

Tao Ding, *Senior Member, IEEE*, Ziyu Zeng, *Student Member, IEEE*, Jiawen Bai, *Student Member, IEEE*, Boyu Qin, *Member, IEEE*, Yongheng Yang, *Senior Member, IEEE*, Mohammad Shahidehpour, *Fellow, IEEE*

***Abstract*—Electric vehicles (EVs) have rapidly developed in recent years and their penetration has also significantly increased, which, however, brings new challenges to power systems. Due to their stochastic behaviors, the improper charging strategies for EVs may violate the voltage security region. To address this problem, an optimal EV charging strategy in a distribution network is proposed to maximize the profit of the distribution system operators while satisfying all the physical constraints. When dealing with the uncertainties from EVs, a Markov decision process (MDP) model is built to characterize the time series of the uncertainties and then the deep deterministic policy gradient based reinforcement learning technique is utilized to analyze the impact of uncertainties on the charging strategy. Finally, numerical results verify the effectiveness of the proposed method.**

***Index Terms*—Electric vehicle, Markov decision process, reinforcement learning, optimal charging strategy**

NOMENCLATURE

Indices and Sets

i,j,n,m	Indices for buses
t,δ	Indices for time periods
w	Index for electric vehicle (EV)
W^{ch}	Set of EVs
L^{net}	Set of branches in a distribution network
B^{net}	Set of buses in a distribution network
$\xi(i)$	Set of branches with the starting node being i
$\Psi(i)$	Set of branches with the end node being i
T^{ch}	Set of time periods
$\mathcal{S}$	Set of states
$\mathcal{A}$	Set of actions
R	Set of rewards
π	Set of policy
J	Set of returns

Parameters

λ^{cha}	EV charging price
$T_{in,w}$ /$T_{out,w}$	Arriving/Leaving time of EV w
$E_{cap,w}$	Battery capacity of EV w
λ_t^{sell}	Retail electricity price for users at time t
λ_t^{buy}	Whole electricity price for the DSO at time t
$P_w^{EV,max}$	Maximum charging and discharging power of EV w
$P_{i,t}^{load}$	Normal load demand on the bus i at time t
$K_{w,i,t}$	Indicator describing whether the w-th EV is plugged in charging station at bus i at time t
SOC_w^{ini}	The initial SOC of EV w
$r_{i,j}$/ $x_{i,j}$	Resistance/ Reactance of branch (i, j)
U_i^{max}/U_i^{min}	Voltage bounds
$I_{i,j}^{max}$	Maximum value of branch current (i, j)
$P_{sub}^{max}/P_{sub}^{min}$	Maximum/minimum of the substation power
SOC_w^{min} / SOC_w^{max}	Maximum/minimum SOC limits of EV w
SOC_w^{exp}	The expected SOC of EV w when leaving
$H_{w,t}$	Charing/discharging state of EV w at time t
S_i	Power capacity of charging station i
$Station_w$	EV w's choice of charging station
$\boldsymbol{W}^{\upsilon}$	Weather conditions at the past time period υ
$\boldsymbol{\Gamma}$	Traffic conditions
$\boldsymbol{q}_{\delta}^{f}$	Prediction for the boundary condition $\boldsymbol{q}_{\delta}$ at time δ
$\boldsymbol{q}$	Boundary conditions
$\boldsymbol{\Gamma}^{\upsilon}$	Traffic information at the past time period υ
$\boldsymbol{T}_{\boldsymbol{in}}^{\upsilon}$ / $\boldsymbol{T}_{\boldsymbol{out}}^{\upsilon}$	T_{in} and T_{out} at the past time period υ
D^{υ}	Load level of the charging stations at the past time period υ

Decision variables

$\bar{P}_{i,t}^{load}$	Integrated load demand with EVs on the bus i at time t
$P_{w,t}^{cha}/P_{w,t}^{dis}$	Charging/Discharging power of EV w at time t
$P_{t,0}$	Power bought from the external grid at time t
$P_{i,t}^{Ccha}/P_{i,t}^{Cdis}$	Integrated charging/discharging power of EVs at bus i and time t
$I_{i,j,t}$	Current on branch (i, j) at time t
$U_{i,t}$	Voltage level at bus i at time t
$P_{i,j,t}/Q_{i,j,t}$	Active/Reactive power flow on branch (i, j) at time t
$SOC_{w,t}$	The state of charge (SOC) of EV w at time t
$\boldsymbol{x}$	Decision variables
s_t	A state at time t
a_t	An action at time t
ω_t	Randomness at time t
r_t	A reward at time t
π_t	A policy at time t
J_t	A return at time t

This work was supported in part by National Key Research and Development Program of China (2016YFB0901900), in part by National Natural Science Foundation of China (Grant 51977166 and U1766215) and in part by China Postdoctoral Science Foundation (2017T100748).

T. Ding, Z. Zeng, J. Bai and B. Qin are with the Department of Electrical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China. T. Ding is the correspondence author (e-mail: tding15@mail.xjtu.edu.cn).

Y. Yang is with the Department of Energy Technology, Aalborg University, Aalborg 9220 Demark.

M. Shahidehpour is with the ECE Department, Illinois Institute of Technology, Chicago, IL 60616 USA.

I. INTRODUCTION

Electric vehicles (EVs) have been developed rapidly and are expected to offer much flexibility to future power systems and especially in the active distribution networks [1]-[3]. In some countries like the U.S., Germany, and China, EV's production has already been industrialized and the construction of infrastructure has been completed. In fact, EVs have many multifarious advantages and technological advances in various aspects. For example, fossil energy consumption in transportation electrification can be significantly reduced, and the greenhouse gas (GHS) emission will be mitigated [4], [5]. From the power system side, the vehicle-to-grid (V2G) mode allows EV

to be served as mobile energy storage that could provide flexibility to the power systems [6]. In a distribution network, the V2G mode could help stabilize the voltage level [7]-[9]. Thus, the EV is able to provide peak shaving services, reduce power system losses, and raise renewable energy penetration rate [10], [11]. The governments in many countries have laid out the blueprint for augmenting the scale of EVs in the future.

However, the integration of EVs in a charging station may lead to a significant voltage violation in a distribution system, especially when the fast charging techniques are performed [12]-[14]. Various studies have been conducted to examine this impact. References 错误!未找到引用源。 and [15] provided a detailed analytical framework to evaluate the impact of plug-in hybrid electric vehicles (PHEV) on distribution systems. It conducted a deterministic analysis to examine distribution system component impact sensitivities, along with the probability based stochastic analysis to depict EV users' behavior. It was revealed that PHEVs could lead to sudden increase on the temperature of electrical components, significant voltage drops, severe unbalance among the three phases, and the increasing power transmission losses. In [17] and [18], simulated scenarios suggested that the voltage drop was related to the penetration and location of EVs in the power network. At certain buses, the voltage level may vary significantly. Moreover, it was concluded that for an uncontrolled charging station, even at relatively modest levels, both the voltage and thermal loading levels may exceed safe operating region. In [19] and [20], the authors studied the impact of different PHEV penetration levels on distribution systems. It pointed out that there were two main factors limiting the EV penetration: one was the sudden increase on the peak of residential load profile, and the other was the temperature of the distribution feeders. Furthermore, it was mentioned that proper charging strategies were necessary to improve the penetration of the EVs. What's more, [21] and [22] found that the distribution transformers may limit the quantity and penetration of the PHEVs.

There are many viable solutions to the above problems and challenges, one of which is to adopt an optimal charging strategy. In [23], a linear programming model for EVs was built to minimize the variance of energy delivered to all the EVs, in which the EVs did not need to provide any service to the network. To alleviate the impacts of EVs on the distribution network, a set of different smart charging scheduling methods were proposed in [24] to flatten sudden peak demands and solve overloading problems by means of the binary particle swarm optimization (BPSO) algorithm. In [25], an optimal charging model and a two-stage margin-based algorithm were discussed to address the high EV penetration considering the additional electricity infrastructure while minimizing the investment expenses. Reference [26] proposed an optimal charging strategy with the network constraints to optimally manage the EVs, so as to provide ancillary services and flexibilities to power systems. Additionally, [27] presented an optimal charging strategy for PHEVs in a DC distribution network considering the battery charging and discharging characteristic. Moreover, an online constrained optimization model was set up to strictly satisfy the under- and over-voltage problems as well as the reverse power flow issues.

It is desired to note that the EV users' uncertain behaviors may significantly affect the optimal charging model for EVs. In order to address this problem, it is necessary to predict EV user's behaviors based on real-time information. Reference [28] used a survey of resident mobility in Spain to find a realistic EV user behavior pattern. However, in practice, such a pattern may vary with time, space, weather, traffic, and other factors. It was then described in [29] that the different EV charging behaviors might affect the load demand as well as the EV charging strategy. Reference [30] further pointed out that difference parking locations would affect the load curve, and in turn changing the optimal charging strategy of EVs. Definitely, many studies focused on this problem and used several different offline methods to describe the uncertain behavior of EV users. The stochastic programming method was used to achieve the optimal management considering the stochastic behavior of the EVs in [31]. Robust optimization can also be used to solve the problem under the uncertain behaviors of EVs. For example, reference [32] gave an optimal charging strategy of EVs in an electricity market by using robust optimization. Reference [33] studied the uncertain behavior of EVs by using hierarchical stochastic predictive control scheme to build a management system for an island microgrid. A hybrid centralized-decentralized charging control scheme was designed to tackle the problem from the massive number of EVs integrated into the power grid [34]. Moreover, in order to deal with the uncertain behavior in an EV routing optimization, a model predictive control-based adaptive scheduling strategy was been proposed in [35]. Besides, a few literatures have proposed various online charging strategies [36]-[39]. The model predictive control (MPC) method was used to describe the uncertainty of EV behaviors, and optimal charging strategies for EVs were proposed accordingly [40]-[42].

However, the EV optimal charging strategy problem is a multi-stage decision making process which can be regarded as a Markov Decision Process (MDP), where the current state is independent of all previous states and actions. Generally, to solve MDP models, an effective solution is the reinforcement learning (RL). Many reinforcement learning algorithms have been proposed and been used in power system [43]. For instance, the Q-learning was developed in 1992 [44], which was a widely used reinforcement learning algorithm. However, the direct Q-learning method was unable to solve large-scale and non-linear problems due to the over-sized problem in the Q Table. To overcome this and enable RL algorithms to be more useful and promising, deep neural networks were implemented in RL algorithms as function approximates. However, the combination of the online RL algorithms with deep neural networks was formerly considered to be difficult and lack of theoretical foundation. Moreover, the neural network-based Q learning methods may suffer the correlation problem among sample data and the serious instability. To address these problems, in 2015, some major problems of the deep reinforcement learning algorithm were successfully overcome and a new artificial agent was developed based on the deep reinforcement learning algorithm, called Deep Q-Network (DQN) [45], [46]. This algorithm utilized an experience replay buffer with an ancillary target neural network, which surpasses the performance of all previous RL algorithms, and it could learn proper policies directly from inputs with high-dimensional sensory data. In recent years,

some researchers observed that the DQN cannot be directly applied to continuous action domains. With the help of the deterministic policy gradient (DPG) algorithm [47] and the experience replay buffer technique, a model-free, off-policy actor-critic-based Deep Deterministic Policy Gradient (DDPG) algorithm was proposed in [48]-[50] by using deep function approximators that can learn policies in high-dimensional and continuous action spaces. DDPG algorithm has already been applied in power system and showed its superiority [51].

This paper proposes a reinforcement learning-based optimal EV charging strategy in a distribution network, maximizing operator profits while addressing grid-level constraints. The contributions of this paper can be summarized as:

(i) A second-order cone programming (SOCP) based optimal EV charging model is proposed for a distribution system operator (DSO), with the consideration of the physical constraints in the distribution network to mitigate the potential voltage issues. To address the uncertain EV user behaviors, an MDP model is built to depict the time series of uncertainties.

(ii) The proposed MDP based EV charging strategy model is solved by the DDPG reinforcement learning (RL) agent, adapt to the uncertainty of EV user behaviors and environmental changes. In this agent, we use an actor network to use observed environment states such as weather and traffic information to give proper actions, and a critical network to evaluate these actions and adapt the actor network to environmental changes.

The rest of this paper is organized as follows. Section II presents a mathematical formulation of the optimal charging problem by a general SOCP model, where the uncertain boundary conditions are molded by the MDP. In Section III, the DDPG-based reinforcement learning agent is employed to solve the proposed optimal vehicle charging strategy model. Numerical results are presented in Section IV to examine the proposed model and method. Finally, Conclusions and further discussions are given in Section V.

II. Formulation of EV Optimal Charging Strategy

In this study, we consider a district distribution system shown in Fig. 1, containing the substation, distribution network, loads and the community-level EV charging stations (e.g., parking lot with charging points). The distributed system operator (DSO) will sell electricity and charging services to users. Due to the limited space area, the maximum number of EVs W^{ch} is a given deterministic parameter. We assume that the DSO can partly control the load demand by performing demand response, and can fully control each EV's charging and discharging power when it is plugged in the charging station. Besides, we assume that each bus is connected to one and only one charging station. Thus, we can use the same index i to represent both bus and charging station. Clearly, the proper charging strategy should be conducted in the context of the existing market price tariff. That means, when the load is high (i.e., voltage level is low), the price is high and vice versa. This is well recognized as the time-of-use price policy in the demand side management. However, the voltage magnitudes in the distribution network may be violated if the EV charging strategy is improper (e.g., charging the EVs at the load peak and discharging the EVs at the load valley). Thus, coordinating the control of demand response and EV charging is critical to the DSO.

Fig. 1. Distribution network and EV charging stations under study.

A. Deterministic EV Optimal Charging Model

An optimization model was set up to investigate an EV optimal charging problem in a distribution system, which is to maximize DSO revenue over one day while satisfying all the physical constraints of the EVs, distribution network, and charging stations. DSO revenue consists of two parts. One is from the profit by buying from the grid and selling electricity to users, i.e. charging revenue. The other comes from the optimal charging services of the charging stations, i.e. load revenue. It should be noted that EV charging rate is only allowed to be at discrete levels due to the current battery and charger technology [52]-[54] which will contribute to the computational complexity of the proposed optimization model. To reduce the problem complexity, many works assumed that future infrastructure improvements will allow continuous charging rates [55]-[57], so we assume that EV charging rate is continuous in this paper. The physical constraints should include power flow, voltage region and the capacity limits of equipment. Finally, the optimization model can be expressed as [58]:

$$\max_{\substack{\overline{P}^{load}_{i,t},P^{cha}_{w,t},P^{dis}_{w,t},P_{t,0},P^{Ccha}_{i,t},P^{Cdis}_{i,t} \\ I_{i,j,t},U_{i,t},P_{i,j,t},Q_{i,j,t},SOC_{w,t}}} \quad profit=$$

$$\underbrace{\lambda^{cha}\sum_{w\in W^{ch}}\left(SOC_{w,T_{out,w}}-SOC_{w,T_{in,w}}\right)E_{cap,w}}_{\text{charging revenue}} + \underbrace{\sum_{t\in T^{ch}}\sum_{i\in B^{net}}\lambda^{sell}_{t}P^{load}_{t,i}-\sum_{t\in T^{ch}}\lambda^{buy}_{t}P_{t,0}}_{\text{load revenue}} \tag{1}$$

$$s.t.\begin{cases}-\overline{P}^{load}_{i,t}=\sum_{m\in\xi(i)}PF_{i,m,t}-\sum_{n\in\psi(i)}(PF_{n,i,t}-r_{n,i}I^2_{n,i,t})\\ -Q^{load}_{i,t}=\sum_{m\in\xi(i)}QF_{i,m,t}-\sum_{n\in\psi(i)}(QF_{n,i,t}-x_{n,i}I^2_{n,i,t})\end{cases},\forall i\in B^{net},t\in T^{ch} \tag{2}$$

$$U^2_{i,t}I^2_{i,j,t}=PF^2_{i,j,t}+QF^2_{i,j,t},\ \forall(i,j)\in L^{net},t\in T^{ch} \tag{3}$$

$$U^2_{i,t}-U^2_{j,t}=2(r_{i,j}PF_{i,j,t}+x_{i,j}QF_{i,j,t})+(r^2_{i,j}+x^2_{i,j})I^2_{i,j,t},\ \forall(i,j)\in L^{net} \tag{4}$$

$$U^{\min}_{i}\le U_{i,t}\le U^{\max}_{i},\quad \forall i\in B^{net},t\in T^{ch} \tag{5}$$

$$0\le I_{i,j,t}\le I^{\max}_{i,j},\quad \forall(i,j)\in L^{net},t\in T^{ch} \tag{6}$$

$$P^{\min}_{sub}\le P^{sub}_{0,t}\le P^{\max}_{sub},t\in T^{ch} \tag{7}$$

$$\overline{P}^{load}_{i,t}=P^{load}_{i,t}-P^{Cdis}_{i,t}+P^{Ccha}_{i,t} \tag{8}$$

$$P_{i,t}^{Cdis} = \sum_{w \in W^{ch}} P_{w,t}^{dis} K_{w,i,t}, \ i \in B^{net}, t \in T^{ch} \tag{9}$$

$$P_{i,t}^{Ccha} = \sum_{w \in W^{ch}} P_{w,t}^{cha} K_{w,i,t}, \ i \in B^{net}, t \in T^{ch} \tag{10}$$

$$SOC_{w,T_{in,w}-1} = SOC_w^{ini}, \ w \in W^{ch} \tag{11}$$

$$SOC_{w,t} = SOC_{w,t-1} + (P_{w,t}^{cha}\eta_{cha} - P_{w,t}^{dis}/\eta_{dis})/E_{cap,w} \ (t \geq 2) \tag{12}$$

$$SOC_{\min} \leq SOC_{w,t} \leq SOC_{\max}, \ w \in W^{ch}, t \in T^{ch} \tag{13}$$

$$SOC_{w,T_{out,w}} \geq SOC_w^{\exp}, \ w \in W^{ch} \tag{14}$$

$$0 \leq P_{w,t}^{cha} \leq H_{w,t} P_w^{EV,\max}, \ w \in W^{ch}, t \in T^{ch} \tag{15}$$

$$0 \leq P_{w,t}^{dis} \leq H_{w,t} P_w^{EV,\max}, \ w \in W^{ch}, t \in T^{ch} \tag{16}$$

$$\sum_{w \in W^{ch}} K_{w,i,t} H_{i,t} \leq S_i, \ \ i \in B^{net}, t \in T^{ch} \tag{17}$$

where $\left(\overline{P}_{i,t}^{load}, P_{w,t}^{cha}, P_{w,t}^{dis}, P_{t,0}, P_{i,t}^{Ccha}, P_{i,t}^{Cdis}, I_{i,j,t}, U_{i,t}, P_{i,j,t}, Q_{i,j,t}, SOC_{w,t}\right)$ are the decision variables of the proposed optimal charging strategy model which are defined in the Nomenclature. The aim of the optimization model is to maximize the profit of the DSO while satisfying all the physical constraints. The physical constraints of the distribution network are depicted in (2)-(8) and the physical constraints of charging station and EVs are formed in (9)-(17). Specifically, equation (2) is the active and reactive power balance constraint at each bus in the distribution network; equation (3) specifies the relationship among the voltage magnitude, current and power; equation (4) characterizes the voltage drop related to the current and power in the whole distribution network; constraints (5)-(7) specify the bound limits for voltage, current and substation; constraint (8) indicates that the aggregated load with the normal load and EVs; constraints (9) and (10) give the integrated discharging and charging powers in the charging station at bus *i*. Equation (11) defines the initial state of charge (SOC) for the *w*-th EV; equation (12) describes the energy storage process corresponding to the charging and discharging powers for the *w*-th EV; equation (13) guarantees the SOC of the *w*-th EV within the allowable region; equation (14) indicates that the SOC for the *w*-th EV should achieve a certain level once the EV leaves at $T_{out,w}$; equations (15) and (16) ensure that the charging and discharging limits should be within the allowable bound during the time when EV is plugged in the charging station. Here, the matrix $H_{w,t}$ in (15) and (16) is employed to describe the charging and discharging state, which can be given as

$$H_{w,t} = \begin{cases} 0 & (t > T_{out,w}, t < T_{in,w}) \\ 1 & (T_{in,w} \leq t \leq T_{out,w}) \end{cases} \quad t \in T^{ch}, w \in W^{ch} \tag{18}$$

Equation (17) guarantees that the capacity of the EV load demand should be limited by the maximum allowable power in the station. Moreover, $K_{w,i}$ is an indicator describing whether the *w*-th EV is plugged in charging station at bus *i*. When $K_{w,i}$=1, the *w*-th EV is plugged in charging station at bus *i*; otherwise, $K_{w,i}$ = 0. However, the optimal charging model is a non-convex optimization model due to the non-convexity of power flow equations, challenging the global optimum. Here, the second-order cones are utilized to relax the power flow equations. At first, a transformation is made for (2)-(4) by replacing the squared voltage magnitude and current as

$$\begin{cases} I_{i,j,t}^2 = F_{i,j,t}, \forall (i,j) \in L^{net} \\ U_{i,t}^2 = V_{i,t}, \ \ \forall i \in B^{net} \end{cases} \tag{19}$$

Thus, the original (2)-(4) will become

$$\begin{cases} P_{i,t}^{sub} - P_{i,t}^{load} = \sum_{m \in \xi(i)} PF_{i,m,t} - \sum_{n \in \pi(i)} (PF_{n,i,t} - r_{n,i} F_{i,j,t}) - P_{i,t}^{Cdis} + P_{i,t}^{Ccha} \\ Q_{i,t}^{sub} - Q_{i,t}^{load} = \sum_{m \in \xi(i)} QF_{i,m,t} - \sum_{n \in \pi(i)} (QF_{n,i,t} - x_{n,i} F_{i,j,t}) \end{cases} \tag{20}$$

$$\forall i \in B^{net}, t \in T^{ch}$$

$$V_{i,t} F_{i,j,t} = PF_{i,j,t}^2 + QF_{i,j,t}^2, \ \ \forall (i,j) \in L^{net}, t \in T^{ch} \tag{21}$$

$$V_{i,t} - V_{j,t} = 2(r_{i,j} PF_{i,j,t} + x_{i,j} QF_{i,j,t}) + (r_{i,j}^2 + x_{i,j}^2) F_{i,j,t}, \ \forall (i,j) \in L^{net} \tag{22}$$

It can be found that only quadratic equalities (21) will lead to the non-convexity, which can be relaxed by conic relaxation, changing equalities into inequalities, such that

$$F_{i,j,t} V_{i,t} \geq PF_{i,j,t}^2 + QF_{i,j,t}^2, \ \ (i,j) \in L^{net}, t \in T^{ch} \tag{23}$$

which is equivalent to

$$\left\| \begin{matrix} 2PF_{i,j,t} \\ 2QF_{i,j,t} \\ F_{i,j,t} - V_{i,t} \end{matrix} \right\|_2 \leq F_{i,j,t} + V_{i,t}, \quad (i,j) \in L^{net}, t \in T^{ch} \tag{24}$$

In addition, when taking the squares for both sides of (5)-(6) to replace ($I_{i,j,t}$, $U_{i,t}$) by ($F_{i,j,t}$, $V_{i,t}$), the optimal charging model can be reformulated as

$$\text{Obj:} \ \ (1) \tag{25}$$

$$s.t. \qquad \left(U_i^{\min}\right)^2 \leq V_{i,t} \leq \left(U_i^{\max}\right)^2 \ \ \forall i \in B^{net}, t \in T^{ch} \tag{26}$$

$$0 \leq F_{i,j,t} \leq \left(I_{i,j}^{\max}\right)^2, \ \ \forall (i,j) \in L^{net}, t \in T^{ch} \tag{27}$$

$$(7)\text{-}(17), (20), (22), (24) \tag{28}$$

B. Markov Decision Process for EV Optimal Charging

It should be noted that the constraints in the optimization model may be affected by the stochastic behavior of EV users. These are usually called boundary conditions, which cannot be revealed precisely in advance and will affect the charging strategy accordingly. Intuitively, these boundary conditions should include each EV's initial SOC, expected SOC, arriving time, leaving time and the choice of charging station. To solve the above optimization model, we must estimate the values of these boundary conditions before solving the model. For convenience, the above model can be compactly written as

$$\max_{\boldsymbol{x}} \quad profit = f(\boldsymbol{x}) \qquad s.t. \quad \boldsymbol{x} \in \Omega(\boldsymbol{q}) \tag{29}$$

where $\Omega(\boldsymbol{q})$ is the feasible region corresponding to the constraints (26)-(28) which is dependent on the boundary conditions $\boldsymbol{q}$, and $\boldsymbol{x}$ is the decision vector made by

$$\boldsymbol{x} = \left\{P_{w,t}^{Cdis}, P_{w,t}^{Ccha}, PF_{i,j,t}, QF_{i,j,t}, F_{i,j,t}, V_{i,t}, SOC_{w,t}\right\} \tag{30}$$

Moreover, the boundary condition vector $\boldsymbol{q}$ is associated with EV user's uncertain behaviors that can be expressed as

$$\boldsymbol{q}=\left(\boldsymbol{T_{in}},\boldsymbol{T_{out}},\boldsymbol{SOC_{ini}},\boldsymbol{SOC_{exp}},\boldsymbol{Station}\right)$$
$$\begin{cases}\boldsymbol{T_{in}}=\left(T_{in,1},...,T_{in,|W^{ch}|}\right),\boldsymbol{T_{out}}=\left(T_{out,1},...,T_{in,|W^{ch}|}\right)\\ \boldsymbol{SOC_{ini}}=\left(SOC_1^{ini},...,SOC_{|W^{ch}|}^{ini}\right),\boldsymbol{SOC_{exp}}=\left(SOC_1^{exp},...,SOC_{|W^{ch}|}^{exp}\right)\\ \boldsymbol{Station}=\left(Station_1,...,Station_{|W^{ch}|}\right)\end{cases}\quad(31)$$

where $|W^{ch}|$ is the cardinality of W^{ch}, denoting the total number of the EVs. $Station_w$ denotes the EV w's choice of charging stations. In fact, the station corresponds to the bus in the distribution network. For example, if $Station_w$=5, it means that the w-th EV will choose the station at bus 5 for charging. As shown in Fig. 2, the boundary conditions are affected by the current environment conditions, which will be changed with the time periods, making the EV user behaviors vary as well. This means that we need to update the predictions at each time step with latest environment observations. In this regard, the boundary conditions can be depicted by time series, such that $(\boldsymbol{q}_0,\boldsymbol{q}_1,...,\boldsymbol{q}_T)$. To characterize the temporal relation of these uncertainties, the MDP framework is adopted to represent the problem as a 5-tuple $\{\mathcal{S},\mathcal{A},R,\pi,J\}$ with state space $\mathcal{S}$=$(s_1,...,s_t,...s_T)$, action space $\mathcal{A}$ =$(a_1,...,a_t,...a_T)$, reward R=$(r_1,...,r_t,...,r_T)$, policy π=$(\pi_1,...,\ \pi_t,...\pi_T)$, and the return J=$(J_1,...,J_t,...J_T)$,. Here, $J_t=E\left[\sum_{\delta=t}^{\infty}\gamma^{\delta}r_{\delta}\right],0<\gamma<1$, where γ is the reward discount factor.

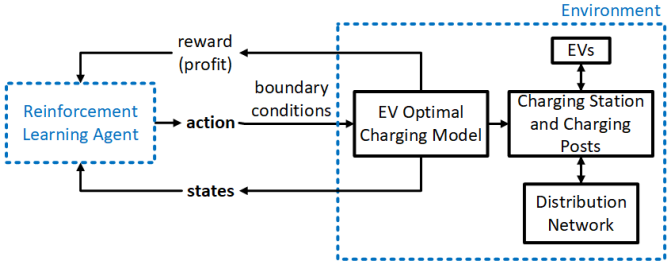

Fig. 2. EV charging control system.

1) State

State is used to describe the environment. As the environment conditions change, the current decisions will be corrected to avoid the violation of the physical constraints with the updated boundary conditions, which will affect the state as well. The observation of environment from the past to the current is defined as a state, which can be written as

$$\boldsymbol{s}_t\in\mathcal{S}=\left(\boldsymbol{W}^{\upsilon},\boldsymbol{\Gamma}^{\upsilon},\boldsymbol{T}_{\boldsymbol{in}}^{\upsilon},\boldsymbol{T}_{\boldsymbol{out}}^{\upsilon},D^{\upsilon},\boldsymbol{d}\right),\upsilon=1,...,t,\quad(32)$$

where $\boldsymbol{W}^{\upsilon}$ refers to the weather information at the past time period υ, including temperature, humidity, cloudiness and air quality index (AQI), which gives

$$\boldsymbol{W}^{\upsilon}=\left(Temperature^{\upsilon},Humidity^{\upsilon},Cloudiness^{\upsilon},AQI^{\upsilon}\right)\quad(33)$$

The weather may affect an EV user's decision on whether to use the EV. For example, if there is a heavy rain in a weekend morning, an EV user may cancel the use plan, leaving the EV in the charging station. $\boldsymbol{\Gamma}^{\upsilon}$ represents the traffic information at the past time period υ which will affect the boundary conditions T_{in} and SOC_{ini}. Generally, for a heavy traffic, the EV users will need more time to arrive and the power consumption will be increased, which thus leads to a large T_{in} and a low SOC_{ini}. Besides, $\boldsymbol{T}_{\boldsymbol{in}}^{\upsilon}$ and $\boldsymbol{T}_{\boldsymbol{out}}^{\upsilon}$ are the T_{in} and T_{out} at the past time period υ, respectively; D^{υ} is the load level of the charging stations at the time period υ and the load level for the i-th charging station is formulated as

$$D_i^{\upsilon}=\frac{\sum_{w\in W^{ch}}K_{w,i,\upsilon}H_{i,\upsilon}}{S_i},\ i\in B^{net},\ \upsilon=1,...,t-1\quad(34)$$

$$K_{w,\arg\min D_i^{\upsilon-1},\upsilon}=\upsilon\,\&\,T_{in,w},\upsilon\in T^{ch},w\in W^{ch},i\in B^{net}\quad(35)$$

where & denotes the logical operator "and".

The charging station load level will also affect the EV user's choice of charging stations. Naturally, an EV user expects to choose a charging station with a relatively low load level because according the current retail price tariff, when the load is high, the price is high and vice versa. This is well recognized as the time-of-use price tariff in the demand side management. The EV user usually expects to find the charging station with the relative low price.

Furthermore, the vector $\boldsymbol{d}$ uses a set of integers to represent dates in a year, where

$$\boldsymbol{d}=\left(season,month,holiday\right)\quad(36)$$

where we use 1 to 4 for *season* to represent spring, summer, autumn, and winter respectively, 1 to 12 for *month* to represent 12 months, and 0 (working day) and 1 (holiday) for *holiday* to represent whether a day is holiday. The three dimensions of the above $\boldsymbol{d}$ may greatly affect the behaviors of EV users. For example, EV user's behaviors on a major holiday and a normal workday are certainly distinct, and the users may act differently in winter and summer.

2) Action

Action is the decision made by the agent to adapt to the environment (i.e., state). In the proposed model, the action $\boldsymbol{a}_t$ at the time t is chosen as the prediction of EV's future behaviors:

$$\boldsymbol{a}_t\in\mathcal{A}=\left\{\boldsymbol{q}_{t+1}^{f},...,\boldsymbol{q}_{\delta}^{f},...,\boldsymbol{q}_{T}^{f}\right\}\quad(37)$$

where $\boldsymbol{q}_{\delta}^{f}$ is the prediction for the boundary condition $\boldsymbol{q}_{\delta}$ at the time period δ=$\{t$+1,…,$T\}$. Note that once an EV is connected to the charging station, some unknown information (T_{in}, SOC_{ini}, SOC_{exp}, *Station*) in $\boldsymbol{q}_{\delta}$ will be revealed, while the leaving time T_{out} remains unknown. This means that (T_{in}, SOC_{ini}, SOC_{exp}, *Station*) in $\boldsymbol{q}_{\delta}$ do not need to forecast anymore in the future.

3) Reward

Any state-action pair is associated with a reward. Considering that the optimal action should maximize the DSO's profits, here we define the reward at the time period t as

$$r_t\left(s_t,a_t\right)=\sum_{i\in B^{net}}\lambda_t^{sell}P_{t,i}^{load}+\lambda^{cha}\sum_{w\in W^{ch}}\left(SOC_{t,w}-SOC_{t-1,w}\right)E_{cap,w}-\sum_{i\in B^{net}}\lambda_t^{buy}P_{t,i}^{sub}\quad(38)$$

4) Policy

A deterministic policy is a map of state-action pairs into real numbers (probability), such that

$$\boldsymbol{\pi}\left(a_t|s_t\right)=\mathbb{P}\left(a_t|s_t\right)\Big|:\mathcal{S}\times\mathcal{A}\rightarrow[0,1]\quad(39)$$

where $\mathbb{P}\left(a_t|s_t\right)$ is the conditional probability under the (s_t,a_t). With the policy function, we are able to make actions based on any environment observations.

5) Return

The objective of the MDP model is to maximize the return with the optimal policy. To find an optimal policy, the action-value function, Q_π: $\mathcal{S}\times\mathcal{A}\rightarrow R$, underlying a policy $\boldsymbol{\pi}$, is defined in (38). Assuming that the first state and the first action are given at the time t, the subsequent actions will be determined by the policy $\boldsymbol{\pi}$.

$$Q_\pi(s_t,a_t)=E_\pi\left[\sum_{\delta=t}^{\infty}\gamma^{\delta-t}r_\delta\,|\,(s_t,a_t)\right] \tag{40}$$

Furthermore, the optimal action-value $Q^*(s_t,a_t)$ at the state-action pair (s_t,a_t) is defined as the maximum of the expected return under (s_t,a_t). In this way, when the exact action-value function is found, an optimal policy $\boldsymbol{\pi}^*$ will naturally be obtained by

$$\boldsymbol{\pi}^*(s_t)=\arg\max_{\boldsymbol{\pi}}\ Q_\pi(s_t,a_t) \tag{41}$$

Furthermore, let the state transition from state s_t to s_{t+1} be $s_{t+1}=\boldsymbol{p}\left(s_{t+1},\left|(s_t,a_t)\right.\right)$, where $\boldsymbol{p}\left(s_{t+1},\left|(s_t,a_t)\right.\right)$ is the transition probability at time t from the state s_t to s_{t+1} under the action a_t, reflecting the environment dynamics. Furthermore, the Bellman equation for the action value function can be formulated as

$$Q_\pi(s_t,a_t)=\sum_{s_{t+1},r_t}p\left(\left(s_{t+1},r_t\right)\left|(s_t,a_t)\right.\right)\left[r_t+\gamma\sum_{a_{t+1}}\pi\left(a_{t+1}\left|s_{t+1}\right.\right)Q_\pi(s_{t+1},a_{t+1})\right] \tag{42}$$

III. Reinforcement Learning Agent for Optimal Vehicle Charging Strategy Model

A. Reinforcement Learning Agent for Markov Decision Process

Traditionally, the MDP model can be solved by the dynamic programming which requires the knowledge of the transition function $\boldsymbol{p}\left(s_{t+1},\left|(s_t,a_t)\right.\right)$. However, this transition function is difficult to obtain, so the precise optimal policy $\boldsymbol{\pi}^*$ and the exact action-vale function in the MDP model are difficult to find. To address this problem, the RL approach is adopted to solve the proposed MDP model. In general, the RL, which is capable of solving MDP problems, has been successfully utilized in dynamic environments, stochastic systems, control systems and etc. For traditional RL methods, finding the optimal action that maximizes $Q(s_t,a_t)$ at each time step needs to enumerate all possible combinations and the computation is very complex. Recently, it was reported in [34] that the DDPG algorithm had excellent advantages in dealing with both continuous states and continuous actions while alleviating the curse of dimensionality. DDPG is a deep reinforcement learning algorithm, leading to an actor-critic architecture, in which the actor produces a current policy and the critic aims to evaluate the current policy.

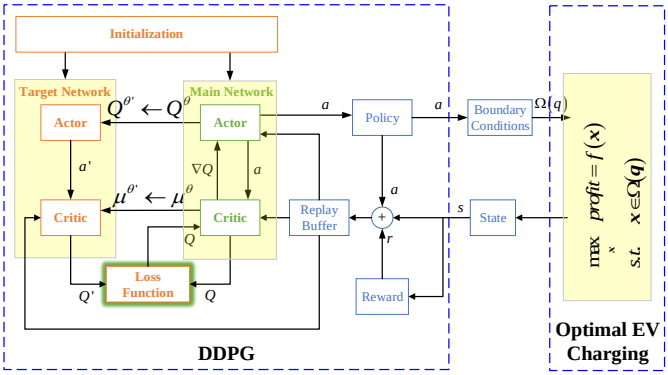

Fig. 3. The structure of the DDPG algorithm.

As shown in Fig.3, the core idea of the DDPG algorithm is to approximate the action-value function $Q(s_t,a_t)$ by an approximator parameterized by $Q(s_t, a_t, \theta^Q)$. This approximator can be updated by the DPG algorithm guarantees the gradient of the off-policy to be equal to the expectation of the gradient of the action-value function. It shows that updating the actor will also use the information from the critic, e.g., the gradient action-value function $\nabla_a Q\left(a_i,s_i,\theta^Q\right)$. At the beginning, we do not know the optimal policy and a stochastic policy is deployed to generate an action a_t. Furthermore, obtaining a reward r_t and a new state s_{t+1} gives a sample $\left\{s_t,a_t,r_t,s_{t+1}\right\}$. Then, a set of samples are collected by $\left\{s_1,a_1,r_1,...,s_t,a_t,r_t,s_{t+1}\right\}$ in the replay buffer to improve the policy and obtain a better reward. In particular, the DDPG algorithm employs an experience replay buffer. Different from the traditional DQN techniques in which the samples are input into the network directly, the samples in DDPG algorithm are stored in a large-size replay buffer instead and then mixed with other stored samples. When the buffer is full, the oldest sample is deleted to make space for the latest sample. At each time step, the RL agent randomly chooses several samples to update the critic. This technique would minimize the correlation among the samples that are used to update the critic. Additionally, directly using the DPG to update $Q(s_t, a_t, \theta^Q)$ may be unstable in many environments. The DDPG algorithm introduces an ancillary target network to improve the stability, where the parameters of the critic in both networks (main network and target network), i.e., $\left(\theta^{\mu'},\theta^{Q'}\right)$ and $\left(\theta^{\mu},\theta^{Q}\right)$ are updated simultaneously.

Finally, the detailed flowchart of the DDPG algorithm can be summarized as follows:

Step 1: Initialize the parameters of the actor-critic structure in the main network with $\left(\theta^{\mu},\theta^{Q}\right)$;

Step 2: Initialize the parameters of the actor-critic structure in the target network with $\left(\theta^{\mu'},\theta^{Q'}\right)\leftarrow\left(\theta^{\mu},\theta^{Q}\right)$;

Step 3: Choose a stochastic policy to generate an action a_t. Furthermore, obtain a reward r_t and a new state s_{t+1}.

Step 4: Add $\left\{s_t,a_t,r_t,s_{t+1}\right\}$ into the replay buffer.

Step 5: Randomly choose N samples in the replay buffer and set the y_o for each sample by

$$y_o=r_o+\gamma Q'\left(s_{o+1},a'_{o+1},\theta^{Q'}\right),o=1,...,N \tag{43}$$

Step 6: Minimize the loss to update the parameters of the critic in the main network by

$$\min_{\theta^Q} \quad L\left(\theta^Q\right)=\frac{1}{N}\sum_{o=1}^{N}\left(Q\left(s_o,a_o,\theta^Q\right)-y_o\right)^2 \tag{44}$$

Step 7: Using stochastic gradient updates the parameters of in the actor in the main network by

$$\nabla_{\theta^\mu}\mu\left(s_i,\theta^\mu\right)\approx\frac{1}{N}\sum_{o=1}^{N}\nabla_{\theta^\mu}\mu\left(a_i,s_i,\theta^\mu\right)\nabla_a Q\left(a_i,s_i,\theta^Q\right) \tag{45}$$

Step 8: Update the parameters of the actor-critic structure in the target network by

$$\begin{aligned}\theta^{\mu'}&\leftarrow\tau\theta^{\mu}+\left(1-\tau\right)\theta^{\mu'}\\ \theta^{Q'}&\leftarrow\tau\theta^{Q}+\left(1-\tau\right)\theta^{Q'}\end{aligned} \tag{46}$$

where τ is soft replacement with $0<\tau<1$.

Step 9: Online application to the optimal EV charging problem and goto **Step 3.**

B. Flowchart of Optimal EV Charging Problem

In this section, we will show the online application of the proposed EV optimal charging model driven by the DDPG algorithm. As shown in Fig. 4, at the beginning, the DSO gathers the boundary conditions $\boldsymbol{q}$, including current and historical information about current weather, traffic condition, and charging stations' load levels, to form the state $\boldsymbol{s}_t$ that will be the input to the RL agent. Then, the proposed DDPG agent will train the RL agent to generate the optimal policy $\boldsymbol{\pi}^*\left(a_t|s_t\right)$. Subsequently, the feasible region $\Omega(\boldsymbol{q})$ can be formed by $\boldsymbol{q}=\boldsymbol{g}\left[\boldsymbol{\pi}^*\left(\boldsymbol{s}\right)\right]$. Based on the optimal policy $\boldsymbol{\pi}^*\left(a_t|s_t\right)$to give an action $\boldsymbol{a}_t$, the corresponding boundary conditions $\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]$ can be estimated and form a feasible region $\Omega\left\{\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]\right\}$, on which we can perform the SOCP to have the optimal charging strategy $\left\{\boldsymbol{x}_1^*,\boldsymbol{x}_2^*,...,\boldsymbol{x}_T^*\right\}$over the next 24 hours by solving the following EV optimal charging problem

$$\begin{aligned}&\max_{\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T} \quad profit=f\left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T\right)\\ &s.t. \quad \left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T\right)\in\Omega\left\{\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]\right\}\end{aligned} \tag{47}$$

Note that for the online closed-loop control, the DSO will conduct the decision at the next hour $\boldsymbol{x}_1^*$ and keep the decisions in the future hours $\left\{\boldsymbol{x}_2^*,...,\boldsymbol{x}_T^*\right\}$. After one hour, we have the true reward r_t, and the environment moves to a new state $\boldsymbol{s}_{t+1}$. Furthermore, we will recalculate the SOCP model under the latest environment observations to update the decisions. This control strategy is called the horizon rolling control. Finally, at the end of each period, the DSO will have all the actual information realized at this hour. Then, all the information will be added in to the historical database for training the boundary conditions at the next hour. At the same time, the RL agent will update the parameters of the DDPG.

Last but not the least, in the above optimization model, we assume that the SOCP model is always feasible under the EV charging strategy. However, the infeasibility of the SOCP may occur due to the uncertain EV behaviors. For example, it is not possible to get all EVs to their expected SOC by their deadlines. At this time, the DSO cannot provide the charging strategy for EVs and the control may be interrupted. In order to prevent this problem, once (47) is infeasible, an unconstrained model is generated by adding penalty functions into the objective function of (47), such that

$$\max_{\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T} \quad profit=f\left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T\right)+\rho\left\|\Omega\left\{\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]\right\}\right\| \tag{48}$$

where ρ is the penalty factor corresponding to the constraints and $\|\bullet\|$ is the regularization norm of the constraint set. Specifically, the constraint set $\Omega\left\{\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]\right\}$ includes both equalities $h_{\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]}\left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T\right)=0$ and inequalities $g_{\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]}\left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T\right)\ge 0$ as

$$\Omega\left\{\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]\right\}=\left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T\right)\left|\begin{cases}g_{\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]}\left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T\right)\ge 0\\ h_{\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]}\left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T\right)=0\end{cases}\right. \tag{49}$$

Furthermore, the unconstrained optimization model can be written as

$$\begin{aligned}\max_{\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T} \quad profit=&f\left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T\right)+\rho_h\left\|h_{\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]}\left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T\right)\right\|_2^2\\ &+\rho_g\left\|\max\left(0,-g_{\boldsymbol{q}\left[\boldsymbol{\pi}^*\left(a_t|s_t\right)\right]}\left(\boldsymbol{x}_1,\boldsymbol{x}_2,...,\boldsymbol{x}_T\right)\right)\right\|_2^2\end{aligned} \tag{50}$$

where ρ_h and ρ_g are the penalty factors corresponding to the equality and inequality constraints, respectively. It can be obvious that solving the unconstrained optimization model (50) is always feasible and the optimal solution can be given if the original model (47) is infeasible.

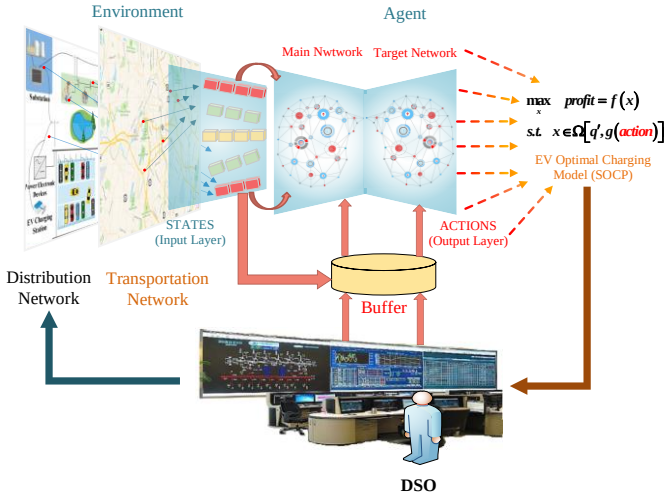

Fig. 4. EV optimal charging strategy with reinforcement learning agent.

IV. Numerical Examples

The proposed MDP model is tested by a distribution network with 33 buses and 3 parking stations as shown in Fig.1. There are 400 EVs in the district and the EV penetration rate is 40%. The detailed information and data for EVs can be available from [59]. We choose one-year EV data as the train set and then use one day data as the test set for online application. The optimal control is conducted over one workday, with each time horizon being one hour. The proposed algorithm is performed on

MATLAB with the Reinforcement Learning Toolbox. We used synthetic data to evaluate proposed method. Moreover, there are two layers in both critic network and actor network. For the critic network, the sizes of the first and second layers are 500 and 300, respectively; learning rate is 10^{-3}; soft replacement is τ=0.001. For the actor network, the first and second layers are 500 and 300, respectively; learning rate is 10^{-4}; soft replacement is τ=0.001. For the DDPG, buffer size is set as 10000, batch size is N=64 and reward discount factor is γ= 0.99.

Table I Tariff TOU price (￥/kWh)

Time (h)	1	2	3	4	5	6	7	8
Tariff price	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.7
Time (h)	9	10	11	12	13	14	15	16
Tariff price	0.7	0.7	0.7	0.7	0.7	0.5	0.5	0.5
Time (h)	17	18	19	20	21	22	23	24
Tariff price	0.5	0.7	0.7	0.7	0.5	0.5	0.5	0.5

For comparison, the traditional offline stochastic programming model is employed. Furthermore, for the online application, we will fix the offline optimal charging strategy, randomly choose 100 scenarios for the boundary conditions and then compute the power flow. To characterize the stochastic nature of the EVs in the district distribution network, the number of EVs for arriving and leaving at different time periods are shown in Fig.5, in which the blue legends represent arriving EVs and the orange legends represent the leaving EVs. In general, the EVs mainly distribute during 4:00-9:00 for leaving and during 16:00-21:00 for arriving. However, the uncertainty during the arriving and leaving time periods is stronger due to the traffic, weather and etc., and the corresponding interval is large. At the other time periods, the uncertainty is small and the corresponding interval is small. For example, at midnight, most of EVs are staying at the charging stations for charging. As a result, the uncertainty is not remarkable.

Fig.6 compares the voltage levels between the proposed and the traditional methods. The proposed method utilizes the reinforce learning technique to solve the MDP model while the traditional method employs the offline stochastic programming to solve the EV charging strategy without considering the MDP model. Moreover, Fig. 7 gives the comparison of the load curves between the proposed and the traditional methods.

Specifically, we can find from the Fig. 6 and Fig. 7 that the MDP model has advantages in guaranteeing the voltage security. On one hand, the traditional load peak usually occurs at 20:00 when the voltage magnitude is at the lowest level for the traditional method. With the optimal charging strategy, the voltage valley will become the voltage peak since the EV discharging will alleviate the heavy load at this moment. On the other hand, the load valley occurs at around 3:00 when the voltage magnitude is at the highest level. The optimal charging strategy will charge EVs at this moment, so that the traditional voltage peak will become the voltage valley. This shows that the optimal charging strategy can benefit the voltage regulation through the charging and discharging at proper time periods.

Moreover, the proposed MDP method is compared with traditional stochastic programming method without MDP and the MPC method. It should be noted that the black solid line represents the voltage curve with MDP model, and the point represents the voltage under 100 possible realizations without MDP model (i.e., by using the offline stochastic model) for the traditional method. Moreover, the blue solid line represents the voltage curve by the MPC method. The results can be found in Fig. 6 that through the online control, the boundary conditions will be updated to the real-time control system. Therefore, the voltage magnitudes controlled by the proposed MDP can always guarantee the voltage security. In contrast, the traditional stochastic programming method performs the offline optimal charging strategy under the forecasted boundary conditions but without the online learning environment for updating the boundary conditions. Thus, the voltage magnitudes may significantly violate the specified boundaries. As for the MPC method, the system will conduct a multi-period optimization model while only deploying the first-period control strategy. Then, the multi-period optimization problem will be recomputed again with the updated forecasted boundary conditions. Although the voltage violation by the MPC method can be alleviated compared to the traditional method, the voltage may still slightly violate the specified boundaries. The reason is that the MPC is an open-loop control framework, and the control for EVs at one time period will affect the boundary conditions of the model at the next time period. In contrast, the online learning system in the proposed method will reconsider the boundary conditions once the current decisions are made and then update the control strategy. This suggests that the boundary conditions and the optimal charging strategy will have strong temporal correlations and the online learning techniques can fully recognize this property, leading to a better solution. In this term, the proposed method can be termed as a closed-loop control framework. However, the traditional stochastic programming and MPC method only focus on the boundary conditions in the view of historical data while neglecting the correlation between the control strategy and the boundary conditions. This may result in the improper control strategy and voltage violation.

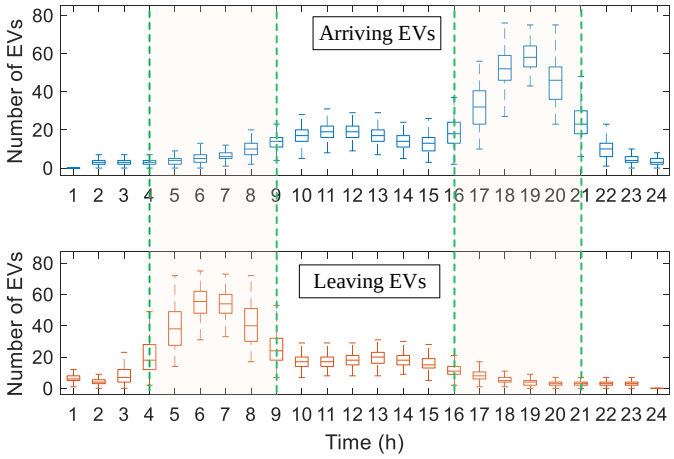

Fig. 5. Number of EVs for arriving and leaving at different time periods.

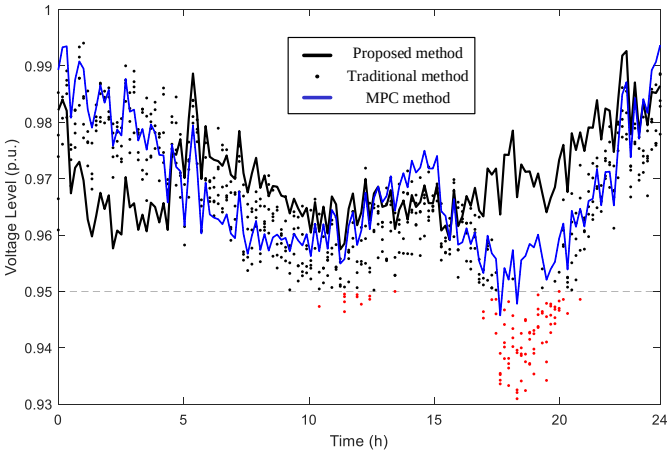

Fig. 6. Comparison of Voltage Magnitudes between the proposed MDP method, the traditional stochastic method without MDP and the MPC method

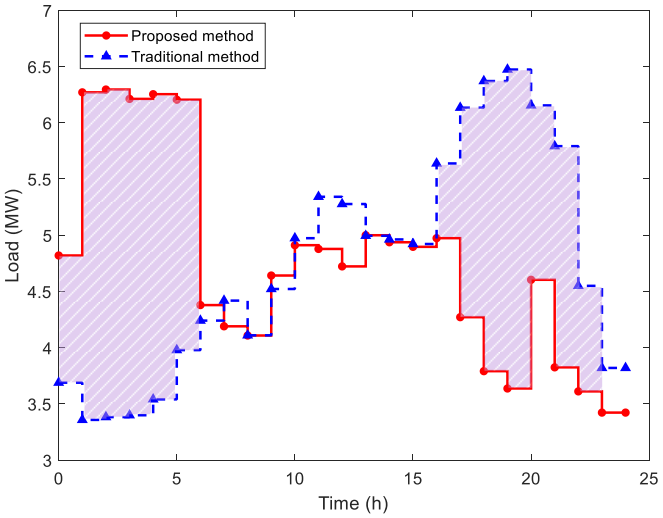

Fig. 7 Comparison of the proposed load curve and traditional load curve.

In addition, the charging and discharging states of three charging stations are depicted in Fig. 8. It is obvious that the time periods for charging is at the load valley, while at load peak for discharging, since the price is high at the load peak and low at the load valley. The charging stations are served for the demand response to reduce the cost. At 18:00~20:00, although the three charging stations are discharging simultaneously to mitigate the under-voltage violation, the discharging powers are coordinated. We randomly choose 20 EVs for illustration and the charging powers are shown in Fig. 9. It can be observed that some EVs at 18:00-20:00 do not participate in the demand response. This is because the EVs for leaving will keep the SOC or the EVs with very small SOC values could not reduce the SOC anymore. Other EVs will shift the charging tasks from 18:00~20:00 to 1:00~5:00 to reduce the cost.

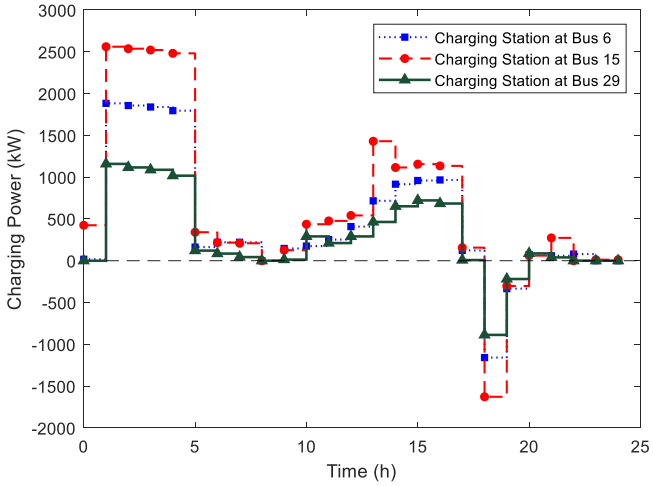

Fig. 8. Charging/discharging curves of charging stations.

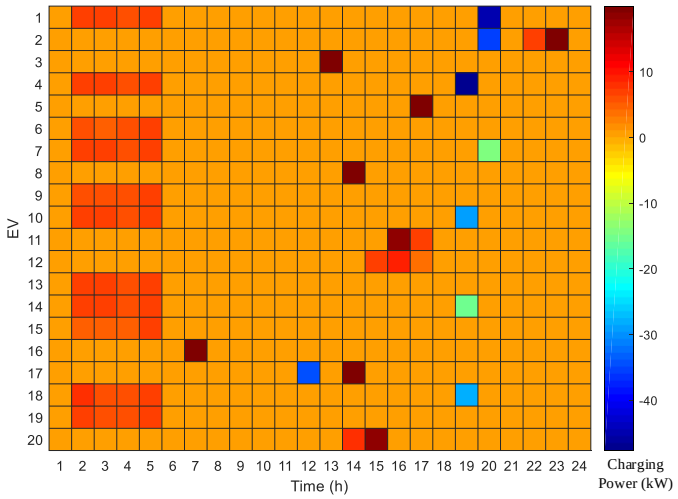

Fig. 9. Charging power of 20 EVs in 24 hours.

Finally, the daily revenue for DSO with two parts (profit for serving load and charging services of EVs) is shown in Fig. 10, where the two parts account for 66% and 34%, respectively. The detailed the hourly revenue data for DSO is shown in Table II, where we can find that the load revenue is high while the charging revenue is low during 18:00-20:00. This is because the electricity price and electricity power consumption during this time period are very high. Therefore, DSO will expect to reduce the charging power of EVs to avoid potential under-voltage problems. Likewise, DSO will expect to increase the charging power of EVs during 2:00-5:00 to maximize its revenue when the load and electricity price are low.

Furthermore, we perform a sensitivity analysis on the EV penetration rate. In addition to the 40% EV penetration rate, we test the model under 10%, 20% and 80% EV penetration rate, respectively. Fig. 11 and Fig. 12 depict the comparisons of voltage levels under different penetrations by the proposed method and traditional method. When the EV penetration rate rises, the fluctuations of voltage magnitudes are severer for both methods. The load peak of EVs is around 17:00 to 20:00, so the voltage levels at charge stations rapidly decline by the traditional method, which will negatively affect the operation of the charging station. In contrast, lots of EVs discharge at the load peak by the proposed method, and the voltage levels do not decline significantly. When the EV penetration rate is pretty high in this

district (e.g. 80%), the voltage levels even rise at the load peak. This is because the cost of buying electricity from the grid during this period is higher than that of electric vehicle discharge. As a result, the EV will discharge the power to the power grid, although there are many EV connects to the charging stations.

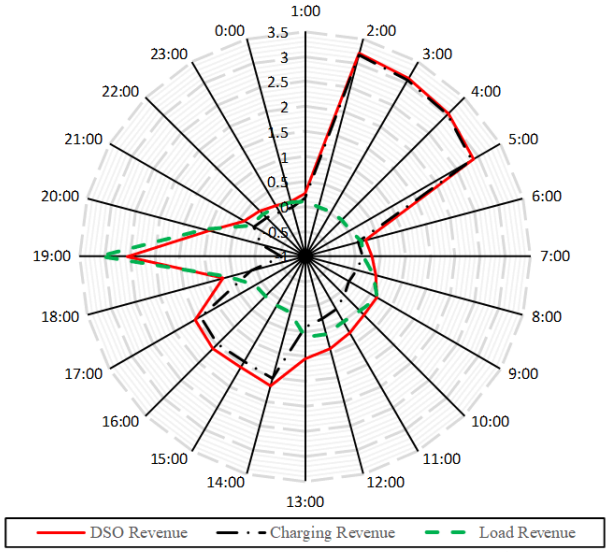

Fig. 10. Analysis of the daily revenue for the DSO (10^3 yuan).

Table III displays the DSO revenue under different penetration level of EVs for the two methods, as well as the computational time. It can be found that for both the two methods, the DSO revenue will increase as the EV penetration level grows. There are two reasons for this phenomenon. First, increasing the penetration level of EVs implies that the charging demand of EV will increase and the first part of the DSO revenue will increase accordingly. Second, higher penetration level of EVs will increase the EV battery capacity and provide more flexibilities to the power grid. When the electricity price is high, the EVs can be fully charged, and the DSO can sell more electricity to grid, suggesting that the second part of the DSO revenue will increase as well. Furthermore, for the same EV penetration level, the DSO revenue by the proposed method is higher than that of the traditional method. In particular, larger difference of the DSO revenue will be obtained between the two methods with the increasing EV penetration level. This shows that the traditional method will sacrifice more benefits under the higher penetration of EVs. However, the computational complexity of the proposed method is higher than the traditional method. Generally, the proposed method will take 50-200 iterations for the reinforcement learning which is a little time-consuming.

Table II. Hourly revenue of the DSO with 40% EV penetration rate

Time (h)	1	2	3	4	5	6
Overall revenue (10^3 ¥)	0.26	3.20	3.11	3.05	2.89	0.26
Charging revenue (10^3 ¥)	0.16	3.17	3.07	3.01	2.84	0.10
Load revenue (10^3 ¥)	0.10	0.03	0.04	0.04	0.05	0.16
Time (h)	7	8	9	10	11	12
Overall revenue (10^3 ¥)	0.33	0.47	0.66	0.65	0.77	0.91
Charging revenue (10^3 ¥)	0.16	0.06	0.00	0.11	0.23	0.30
Load revenue (10^3 ¥)	0.17	0.41	0.66	0.54	0.52	0.61
Time (h)	13	14	15	16	17	18
Overall revenue (10^3 ¥)	1.04	1.67	1.56	1.61	1.54	0.69
Charging revenue (10^3 ¥)	0.42	1.52	1.42	1.48	1.41	0.11
Load revenue (10^3 ¥)	0.62	0.15	0.14	0.13	0.14	0.58
Time (h)	19	20	21	22	23	24
Overall revenue (10^3 ¥)	2.55	0.98	0.42	0.28	0.17	0.13
Charging revenue (10^3 ¥)	0.54	0.09	0.17	0.09	0.02	0.01
Load revenue (10^3 ¥)	3.09	1.07	0.25	0.19	0.15	0.12

Table III. DSO revenue and computational complexity under different EV penetration levels for two methods

Method	EV Penetration level (%)	Revenue (10^3 yuan/day)	Computational Time (s)
Proposed Model	10	9.494	38.11
	20	17.943	98.45
	40	29.173	212.34
	80	50.229	416.54
Traditional Model	10	8.726	3.19
	20	13.854	5.65
	40	20.354	6.35
	80	32.379	14.72

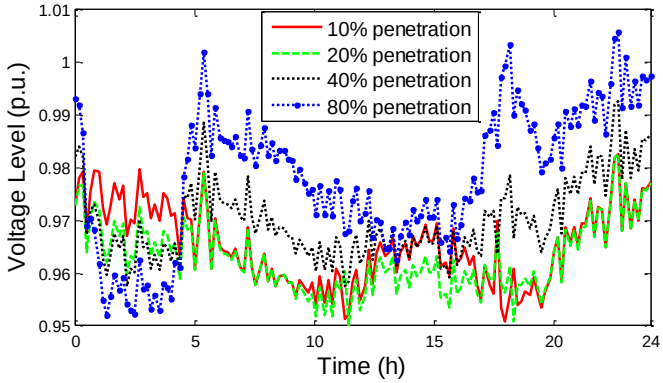

Fig. 11. Comparisons of voltage levels under different penetrations. (The proposed method)

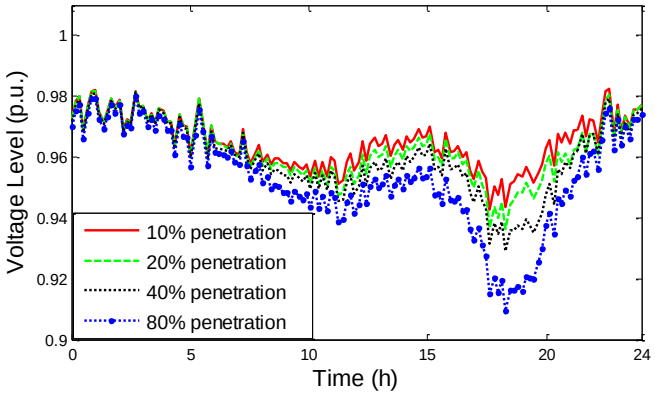

Fig. 12. Comparisons of voltage levels under different penetrations. (The traditional method)

V. Conclusions

This paper proposes a reinforcement learning-based optimal charging strategy model for a DSO to address the voltage violation problem. In the context of uncertain EV users' behaviors, an MDP framework is utilized for the temporal properties of the uncertainties. Furthermore, the deep deterministic policy gradient algorithm is employed to solve the proposed model. Simulation results suggest that the optimal charging strategy for EVs is served as a kind of demand response which will be beneficial for the voltage regulation through the charging and discharging at proper time periods. Moreover, it verifies that the proposed method can strictly guarantee the voltage security while the traditional stochastic approach cannot. This is because the control for EVs at one time period will affect the boundary conditions

of the model at the next time period. The MDP and the online learning techniques can fully take the temporal correlations into account.

REFERENCES

[1] C. Liu, K. T. Chau, D. Wu and S. Gao, "Opportunities and challenges of Vehicle-to-Home, Vehicle-to-Vehicle, and Vehicle-to-Grid technologies," *Proc. IEEE*, vol. 101, no. 11, pp. 2409-2427, Nov. 2013.

[2] J. Zhang, Y. Pei, J, Shen, " Charging Strategy Unifying Spatial-Temporal Coordination of Electric Vehicles," *IET Gener. Transm. Distrib.*, Early Acess, 2020.

[3] T. Ding, J. Bai, P. Du, B. Qin and et al, "Rectangle Packing Problem for Battery Charging Dispatch Considering Uninterrupted Discrete Charging Rate," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 2472-2475, 2019.

[4] J. Zhang, X. Sun, L. Jia and Y. Zhou, "Electric passenger vehicles sales and carbon dioxide emission reduction potential in China's leading markets," *Jour. Clean. Prod.*, vol. 243, 118607, Jan. 2020.

[5] Y. Sun, Z. Chen, Z. Li, W. Tian and M. Shahidehpour, "EV Charging Schedule in Coupled Constrained Networks of Transportation and Power System," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4706-4716, Sept. 2019.

[6] D. Papadaskalopoulos, G. Strbac, P. Mancarella, M. Aunedi and V. Stanojevic, "Decentralized Participation of Flexible Demand in Electricity Markets—Part II: Application With Electric Vehicles and Heat Pump Systems," *IEEE Trans. Power Syst*., vol. 28, no. 4, pp. 3667-3674, Nov. 2013.

[7] D. He, S. Chan and M. Guizani, "Privacy-friendly and efficient secure communication framework for V2G networks," *IET Communications*, vol. 12, no. 3, pp. 304-309, 20 2 2018.

[8] Q. Sun, H. Lv, S. Gao, K. Wei and M. Mauersberger, "Optimized Control of Reversible VSC with Stability Mechanism Study in SMES Based V2G System," *IEEE Transactions on Applied Superconductivity*, vol. 29, no. 2, pp. 1-6, March 2019, Art no. 5401106.

[9] U. B. Baloglu and Y. Demir, "Economic analysis of hybrid renewable energy systems with V2g integration considering battery life," *Energy Procedia*, vol. 107, pp. 242-247, Feb. 2017.

[10] A. T. Eseye, M. Lehtonen, T. Tukia, S. Uimonen and R. J. Millar, "Optimal Energy Trading for Renewable Energy Integrated Building Microgrids Containing Electric Vehicles and Energy Storage Batteries," *IEEE Access*, vol. 7, pp. 106092-106101, 2019.

[11] W. Kempton, J. Tomić, "Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy," *Journal of Power Sources*, vol. 144, no. 1, pp. 280-294, Jun. 2005.

[12] S. Singh, S. Jagota and M. Singh, "Energy management and voltage stabilization in an islanded microgrid through an electric vehicle charging station," *Sust. Cities Soc*., vol. 41, pp. 679-694, Aug. 2018.

[13] S. Rezaee, E. Farjah and B. Khorramdel, "Probabilistic analysis of plug-in electric vehicles impact on electrical grid through homes and parking lots," *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 1024-1033, Oct. 2013.

[14] F. Marra et al., "Improvement of local voltage in feeders with photovoltaic using electric vehicles," *IEEE Trans. Power Syst*., vol. 28, no. 3, pp. 3515-3516, Aug. 2013.

[15] J. Zhang, Y. Pei, J. Shen and et al, "Optimal charging strategy for electric vehicles using symbolic–graphic combination principle," *IET Gener. Transm. Distrib*., vol. 13, no. 13, pp. 2761-2769, 2019.

[16] A. Rezaei, J. B. Burl, M. Rezaei and B. Zhou, "Catch Energy saving opportunity in charge-depletion mode, a real-time controller for plug-in hybrid electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11234-11237, Nov. 2018.

[17] P. Richardson, D. Flynn and A. Keane, "Impact assessment of varying penetrations of electric vehicles on low voltage distribution systems," *2010 IEEE PES General Meeting*, Providence, RI, 2010, pp. 1-6.

[18] S. Tao, K. Liao, X. Xiao, J. Wen, Y. Yang and J. Zhang, "Charging demand for electric vehicle based on stochastic analysis of trip chain," *IET Gener. Transm. Distrib.*, vol. 10, no. 11, pp. 2689-2698, Aug. 2016.

[19] K. Schneider, C. Gerkensmeyer, M. Kintner-Meyer and R. Fletcher, "Impact assessment of plug-in hybrid vehicles on Pacific Northwest distribution systems", *IEEE PES General Meeting-Conversion and Delivery of Electrical Energy*, Pittsburgh, PA, 2008, pp. 1-6.

[20] H. Lin, Y. Liu, Q. Sun, R. Xiong, H. Li and R. Wennersten, "The impact of electric vehicle penetration and charging patterns on the management of energy hub – A multi-agent system simulation," *Appl. Energy*, vol. 230, pp. 189-206, Nov. 2018,

[21] C. Gerkensmeyer, M. Kintner-Meyer and J. G. DeSteese, Technical Challenges of Plug-in Hybrid Electric Vehicles and Impacts to the US Power System: Distribution System Analysis, Pacific Northwest National Laboratory Report, 2010. [online]. Available: http://www.pnl.gov/main/publications/external/technical_reports/PNNL-19165.pdf.

[22] M. K. Gray and W. G. Morsi, "Power quality assessment in distribution systems embedded with plug-in hybrid and battery electric vehicles," *IEEE Trans. Power Syst*., vol. 30, no. 2, pp. 663-671, March 2015.

[23] J. Villalobos, I. Zamora, K. Knezović and M. Marinelli, "Multi-objective optimization control of plug-in electric vehicles in low voltage distribution networks," *Appl. Energy*, vol. 180, pp. 155-168, 2016.

[24] P. Richardson, D. Flynn and A. Keane, "Optimal charging of electric vehicles in low-voltage distribution systems," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 268-279, Feb. 2012.

[25] F. Baccino, S. Grillo, S. Massucco and F. Silvestro, "A two-stage margin-based algorithm for optimal plug-in electric vehicles scheduling," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 759-766, March 2015.

[26] E. L. Karfopoulos and N. D. Hatziargyriou, "A multi-agent system for controlled charging of a large population of electric vehicles," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1196-1204, May 2013.

[27] M. Tabari and A. Yazdani, "An energy management strategy for a dc distribution system for power system integration of plug-in electric vehicles," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 659-668, March 2016.

[28] M. Alonso, H. Amaris, J. Germain and J. Galan, "Optimal Charging Scheduling of Electric Vehicles in Smart Grids by Heuristic Algorithms," *Energies*, vol. 7, no. 4, pp.2449-2475, Apr. 2014.

[29] D. Steen, L. A. Tuan, O. Carlson and L. Bertling, "Assessment of electric vehicle charging scenarios based on demographical data," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1457-1468, Sept. 2012.

[30] D. Fischer, A. Harbrecht, A. Surmann and R. McKenna, "Electric vehicles' impacts on residential electric local profiles – A stochastic modelling approach considering socio-economic, behavioural and spatial factors," *Appl. Energy*, vol. 233–234, pp. 644-658, Jan. 2019.

[31] P. Sánchez-Martín, S. Lumbreras and A. Alberdi-Alén, "Stochastic Programming Applied to EV Charging Points for Energy and Reserve Service Markets," *IEEE Trans. Power Syst.*, vol. 31, no. 1, pp. 198-205, Jan. 2016.

[32] N. Korolko and Z. Sahinoglu, "Robust Optimization of EV Charging Schedules in Unregulated Electricity Markets," in IEEE Transactions on Smart Grid, vol. 8, no. 1, pp. 149-157, Jan. 2017.

[33] Y. Zou, Y. Dong, S. Li and Y. Niu, "Multi-time hierarchical stochastic predictive control for energy management of an island microgrid with plug-in electric vehicles," *IET Gener. Transm. Distrib.*, vol. 13, no. 10, pp. 1794-1801, 21 5 2019.

[34] R. Wang, G. Xiao and P. Wang, "Hybrid Centralized-Decentralized (HCD) Charging Control of Electric Vehicles," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6728-6741, Aug. 2017.

[35] C. Li, T. Ding, X. Liu and C. Huang, "An Electric Vehicle Routing Optimization Model With Hybrid Plug-In and Wireless Charging Systems," *IEEE Access*, vol. 6, pp. 27569-27578, 2018.

[36] A. Ghosh and V. Aggarwal, "Control of Charging of Electric Vehicles through Menu-Based Pricing". *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5918-5929, Nov. 2018.

[37] T. Nathaniel, B. Turan, and M. Alizadeh . "Online Charge Scheduling for Electric Vehicles in Autonomous Mobility on Demand Fleets.",2019.

[38] H. Zhang, Z. Hu, Z. Xu and Y. Song, "Optimal Planning of PEV Charging Station With Single Output Multiple Cables Charging Spots," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2119-2128, Sept. 2017.

[39] R. Valentin, et al. "An Online Mechanism for Multi-Unit Demand and its Application to Plug-in Hybrid Electric Vehicle Charging.", *J. Artif. Intell. Res.* vol.48, no.1 pp.175-2304, 2013

[40] B. Wang, Y. Wang, H. Nazaripouya, C. Qiu, C. Chu and R. Gadh, "Predictive Scheduling Framework for Electric Vehicles With Uncertainties of User Behaviors," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 52-63, Feb. 2017.
[41] Z. J. Lee et al., "Large-Scale Adaptive Electric Vehicle Charging," *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Anaheim, CA, USA, 2018, pp. 863-864.
[42] Yi Guo, Jingwei Xiong, Shengyao Xu and Wencong Su, "Two-stage economic operation of microgrid-like electric vehicle parking deck," *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, Dallas, TX, 2016, pp. 1-1.
[43] Vázquez-Canteli, José R., and Zoltán Nagy. "Reinforcement learning for demand response: A review of algorithms and modeling techniques." *Applied energy*, 235 (2019): 1072-1089.
[44] Christopher J. C. H. Watkins and Peter Dayan, "Q-Learning," *Mach. Learn.*, vol. 8, no. 3-4, pp.279-292, 1992.
[45] V. Mnih, A. P. Badia, M. Mirza *et al*, "Asynchronous Methods for Deep Reinforcement Learning," *Proc. 33rd Int. Conf. Mach. Learn*, vol. 48, 2016, pp. 1928–1937.
[46] V. Mnih *et al*, "Human-level control through deep reinforcement learning", *Nature*, vol. 518, pp. 529-533, Feb. 2015.
[47] D. Silver, G. Lever, N. Heess, *et al*. "Deterministic policy gradient algorithms," *31st International Conference on Machine Learning*, 2014.
[48] T. P. Lillicrap, J. J. Hunt, A. Pritzel et al, "Continuous control with deep reinforcement learning," *International Conference on Learning Representations*, 2016.
[49] Barth-Maron, G., Hoffman, M. W., Budden, D., Dabney, W., Horgan, D., TB, D., Muldal, A., Heess, N., and Lillicrap, T. "Distributional policy gradients," *International Conference on Learning Representations*, 2018.
[50] Liu, Tao, et al. "A novel deep reinforcement learning based methodology for short-term HVAC system energy consumption prediction," *International Journal of Refrigeration*, 107 (2019): 39-51.
[51] Y. Ye, D. Qiu, M. Sun, D. Papadaskalopoulos and G. Strbac, "Deep Reinforcement Learning for Strategic Bidding in Electricity Markets," *IEEE Trans. Smart Grid*, early access.
[52] L. Gan, U. Topcu, and S. H. Low, "Stochastic distributed protocol for electric vehicle charging with discrete charging rate," *Proc. IEEE Power Energy Soc. Gen. Meet.*, San Diego, CA, USA, pp. 1–8, 2012.
[53] G. Binetti, A. Davoudi, D. Naso, B. Turchiano, and F. L. Lewis, "Scalable real-time electric vehicles charging with discrete charging rates," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2211–2220, Sep. 2015.
[54] B. Sun, Z. Huang, X. Tan, and D. H. Tsang, "Optimal scheduling for electric vehicle charging with discrete charging levels in distribution grid," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 624–634, Mar. 2018.
[55] L. Cheng, Y. Chang, Q. Wu, W. Lin and C. Singh, "Evaluating Charging Service Reliability for Plug-In EVs From the Distribution Network Aspect," *IEEE Trans. Sustain. Energy*, vol. 5, no. 4, pp. 1287-1296, Oct. 2014.
[56] Y. Yang, Q. Jia, X. Guan, X. Zhang, Z. Qiu and G. Deconinck, "Decentralized EV-Based Charging Optimization With Building Integrated Wind Energy," I*EEE Transactions on Automation Science and Engineering*, vol. 16, no. 3, pp. 1002-1017, July 2019.
[57] Y. Zheng, Y. Song, D. J. Hill and K. Meng, "Online Distributed MPC-Based Optimal Scheduling for EV Charging Stations in Distribution Systems," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 2, pp. 638-649, Feb. 2019.
[58] P. Li, X. Duan, J. Yang, Q. Zhang, Y. Zhao and Z. Hu, "Coordinated EV charging and reactive power optimization for radial distribution network using mixed integer second-order cone programming," *ITEC Asia-Pacific*, Harbin, pp. 1-5, 2017.
[59] https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

Tao Ding (SM'19) received the B.S.E.E. and M.S.E.E. degrees from Southeast University, Nanjing, China, in 2009 and 2012, respectively, and the Ph.D. degree from Tsinghua University, Beijing, China, in 2015. During 2013 and 2014, he was a Visiting Scholar in the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA. He is currently an Associate Professor in the State Key Laboratory of Electrical Insulation and Power Equipment, the School of Electrical Engineering, Xi'an Jiaotong University. His current research interests include electricity markets, power system economics and optimization methods, and power system planning and reliability evaluation. He has published more than 60 technical papers and authored by "Springer Theses" recognizing outstanding Ph.D. research around the world and across the physical sciences—*Power System Operation with Large Scale Stochastic Wind Power Integration*. He received the excellent master and doctoral dissertation from Southeast University and Tsinghua University, respectively, and Outstanding Graduate Award of Beijing City. Dr. Ding is an Associate Editor of CSEE Journal of Power and Energy Systems.

Ziyu Zeng (S'20) received the B.S. degree from the School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China, in 2018. He is currently working toward the M.S. degree at Xi'an Jiaotong University. His major research interests include power system optimization and renewable energy integration.

Jiawen Bai (S'20) received the B.S. degree from the School of Electrical Engineering, Tsinghua University, Beijing, China, in 2018. He is currently working toward the Ph. D. degree at Xi'an Jiaotong University. His major research interests include power system economic dispatch and renewable energy characteristic analysis.

Boyu Qin (**M'18**) received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, Shaanxi, China, the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2011 and 2016 respectively. He is currently a lecturer at the school of electrical engineering, Xi'an Jiaotong University. His research interests include power system stability analysis and control, input-to-state stability theory, renewable energy generation, and smart grids.

Yongheng Yang (SM'17) received the B.Eng. degree in electrical engineering and automation from Northwestern Polytechnical University, Shaanxi, China, in 2009 and the Ph.D. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 2014.
He was a postgraduate student with Southeast University, China, from 2009 to 2011. In 2013, he spent three months as a Visiting Scholar at Texas A&M University, USA. Currently, he is an Associate Professor with the Department of Energy Technology, Aalborg University, where he also serves as the Vice Program Leader for the research program on photovoltaic systems. His current research is on the integration of grid-friendly photovoltaic systems with an emphasis on the power electronics converter design, control, and reliability.
Dr. Yang is the Chair of the IEEE Denmark Section. He serves as an Associate Editor for several prestigious journals, including the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, the IEEE TRANSACTIONS ON POWER ELECTRONICS, and the IEEE Industry Applications Society (IAS) Publications. He is a Subject Editor of the IET Renewable Power Generation for Solar Photovoltaic Systems, including the Maximum Power Point Tracking. He was the recipient of the 2018 IET Renewable Power Generation Premium Award and was an Outstanding Reviewer for the IEEE TRANSACTIONS ON POWER ELECTRONICS in 2018.

Mohammad Shahidehpour (F'01) received an Honorary Doctorate degree from the Polytechnic University of Bucharest, Bucharest, Romania, in 2009. He is a University Distinguished Professor, and Bodine Chair Professor and Director of the Robert W. Galvin Center for Electricity Innovation at Illinois Institute of Technology. He is a member of the US National Academy of Engineering, Fellow of IEEE, Fellow of the American Association for the Advancement of Science (AAAS), and Fellow of the National Academy of Inventors (NAI).